Accommodating measurement scale uncertainty in extreme value analysis of northern north sea storm severity

David Randell, Philip Jonathan, Shell Technology Centre Thornton, UK & **Kevin Ewans,** Shell International Exploration and Production & **Daniel Reeve,** University of Lancaster, UK

Motivation: extreme waves

- Modelling storm severity is critical to the design and reliable operation of marine structures.
- The extreme value analysis of significant wave heights, denoted H_S is one of many cases where data relating to the same physical process may be measured on more than one scale; for example $H_S{}^2$ is proportional to the drag force induced by the waves on a structure and may be of interest.
- Return value estimates obtained from the square root of an extreme value model fitted to $H_S{}^2$ data differ from estimates fitted to H_S data, since different tail behaviour is indicated by the two sets of parameter estimates.

Data

- The data examined are significant wave heights (H_S) from a northern North Sea hindcast for the period 1st October 1964 to 31st March 1995 inclusive sampled continuously at 3h intervals for each of 50 locations.
- Significant wave height (H_S) is calculated as four times the standard deviation of displacement from mean sea level and is a measure of ocean energy .
- Certain months and directions were omitted to try to create a more homogenous data set. Firstly, the six month period from 31st March to 1st October was omitted from each year since storms from winter months are of greater magnitude. Secondly, directional effects were accounted for by omitting data originating from 150-350° which were of a lower magnitude (shown in figure 1 below).
- With covariate effects have been removed we have ≈3000 observations per location.

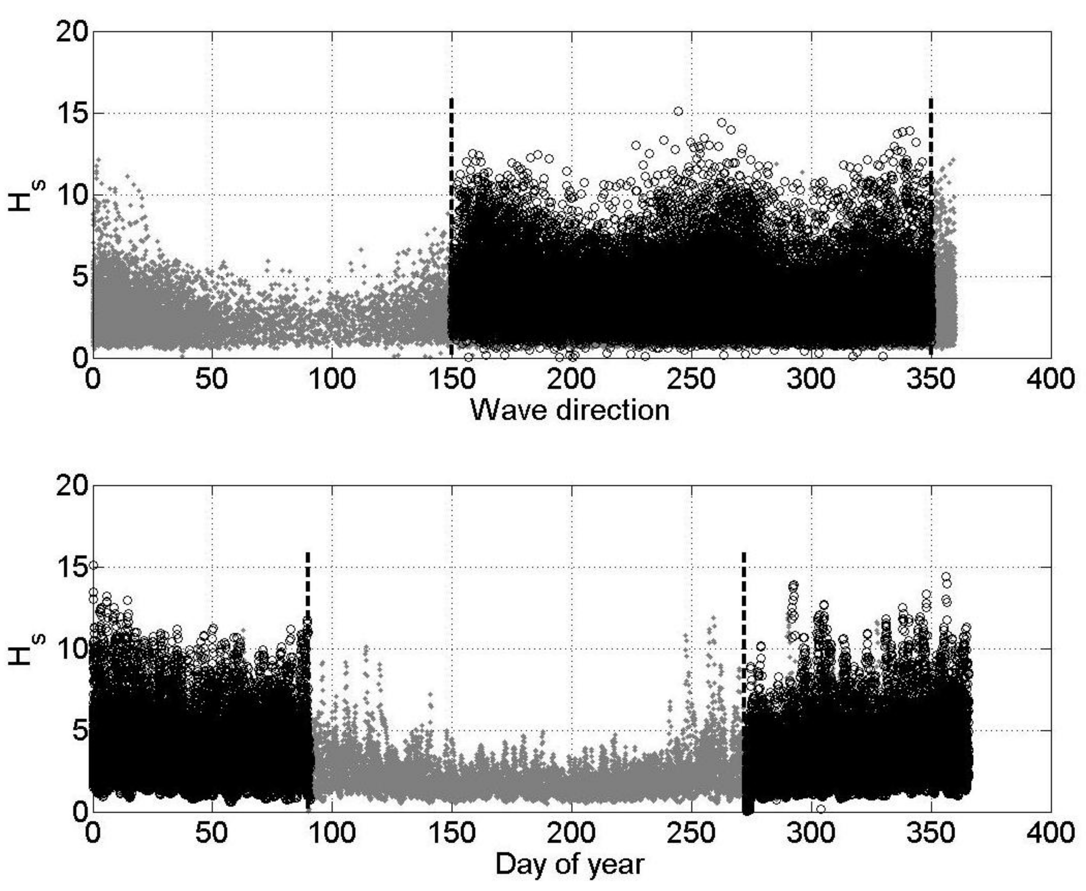

Figure1: Significant wave heights as a function of covariate effects of wave direction and day of the year. Wave directions of 150-350° and the months from October to March are shown in black circles. Restricting to these directions and months results in a data set which is broadly homogeneous.

Extreme Value Analysis

- Extreme value analysis uses sample data from rare events and via a process of extrapolation attempts to make rational predictions about the probable outcome of future events (Coles 2001).

The three parameter generalised Extreme Value Distribution (GEV) has cumulative distribution function:

$$G(x)=\exp\left[-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}_{+}^{-1/\xi}\right];\quad \sigma>0;\quad \mu,\xi\in\mathbb{R};$$

where μ,σ and ξ are the location, shape and scale parameters respectively.

For a data set divided into block maxima, the GEV is used to model the distribution of the maximum values of each block (e.g. the weekly maxima of 3 hourly H_s values) . Define:

$$M_{X,n}=\max\{X_1,...,X_n\},$$

then as in Wadsworth et al. (2010) define:

$$M_{Y,n}=\frac{M_{X,n}^{\lambda}-1}{\lambda}\sim \mathrm{GEV}(\mu_Y,\sigma_Y,\xi_Y)$$

incorporating scale selection into the model through a fourth parameter, λ, using a Box-Cox transformation (Box and Cox, 1964).

Under block maxima the model become a Non homogeneous Poisson process (NHPP) with intensity.

$$\frac{1}{\sigma_y}\left[1+\xi_y\left(\frac{x-\mu_y}{\sigma_y}\right)\right]^{\frac{-1}{\xi_y}-1}$$

- The scale parameter λ means that uncertainty in scale is automatically included in the modelling

Markov Chain Monte Carlo

- Analysis is carried out using MCMC employing a variation of the Metropolis Hastings Random Walk Algorithm (Hastings, 1970).
- All 4 parameters (μ,σ,ξ and λ) were updated in the MCMC
- An Adaptive Metropolis Hastings model was adopted whereby the step variance of each parameter was allowed to vary if the acceptance rate fell outside 25%-50% to ensure good mixing.

Analysis of Northern North Sea Data

- The data are fitted using the both the 3P (without scale parameter) and 4P (with scale parameter) point process models
- Detailed results of parameter estimates and return levels are provided for locations T (top of area), C (central location) and B (bottom of area) in Figure 2 and 3 below.
- Return level estimates are consider denoted $Q_{0.1}$, Q_1, Q_{10} and Q_{100} respectively. Where the Q_i event represent the i times the length of the data.

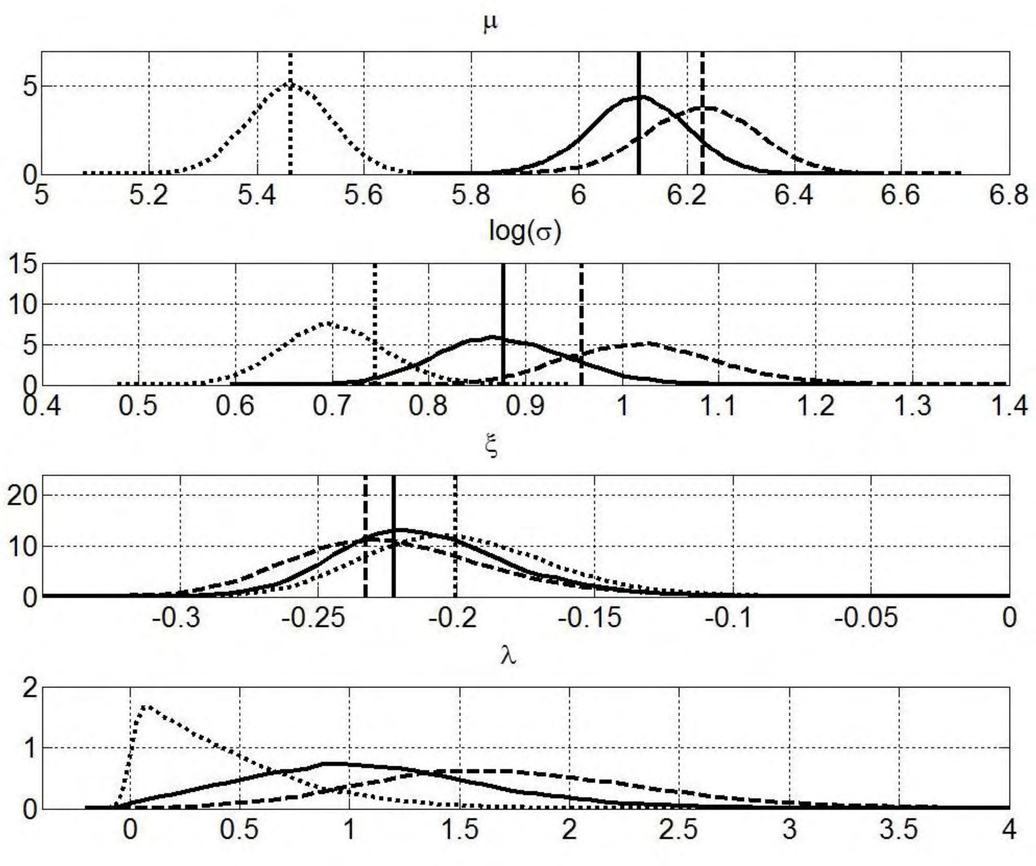

Figure2: Posterior distributions for 4P model parameters at location T (top, dashed line), location C (centre, solid line) and location B (bottom, dotted line). 3P maximum likelihood estimates for, μ log(σ) and ξ re also indicated as vertical lines for comparison. (Results are for 75% threshold and 48 hour blocked data)

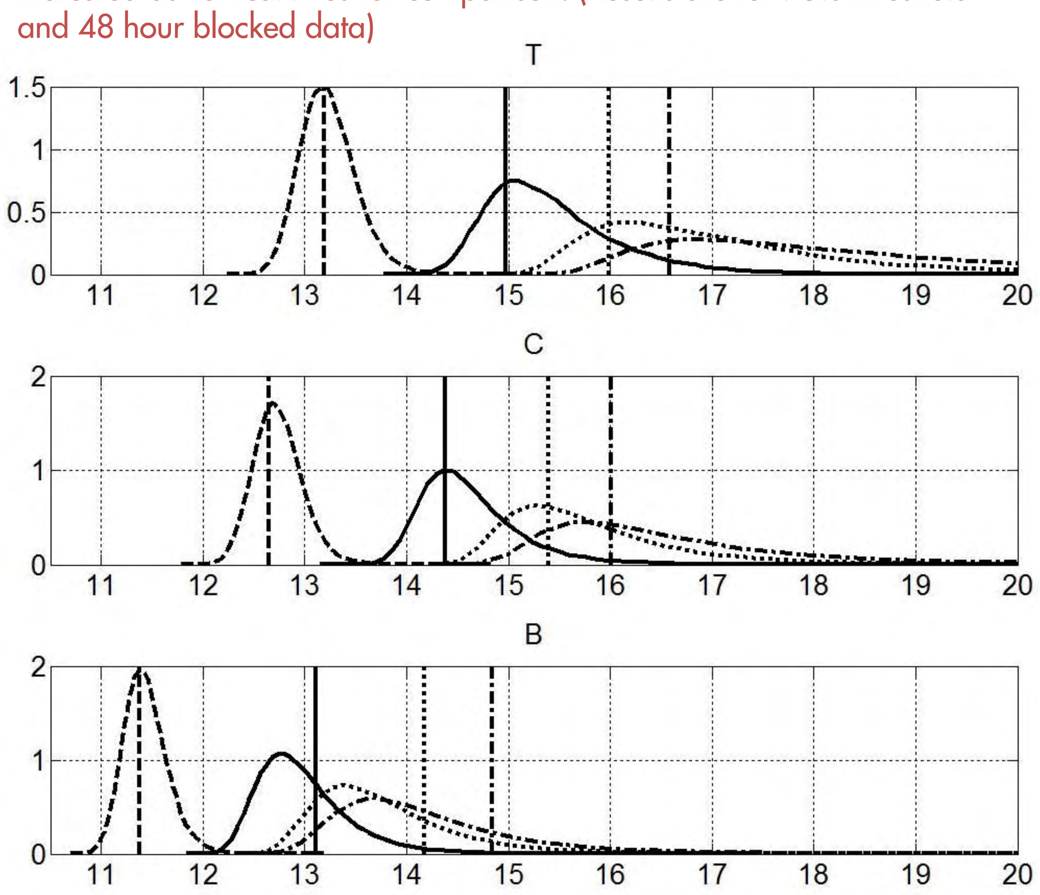

Figure3: Posterior distributions for $Q_{0.1}$ (dashed line), Q_1 (solid line), Q_{10} (dotted line) and Q_{100} (dot-dashed line) return values for 4P model at locations T (top), C (centre) and B (bottom). 3P model maximum likelihood estimates are also shown as vertical lines for comparison. (Results are for 75% threshold and 48 hour blocked data)

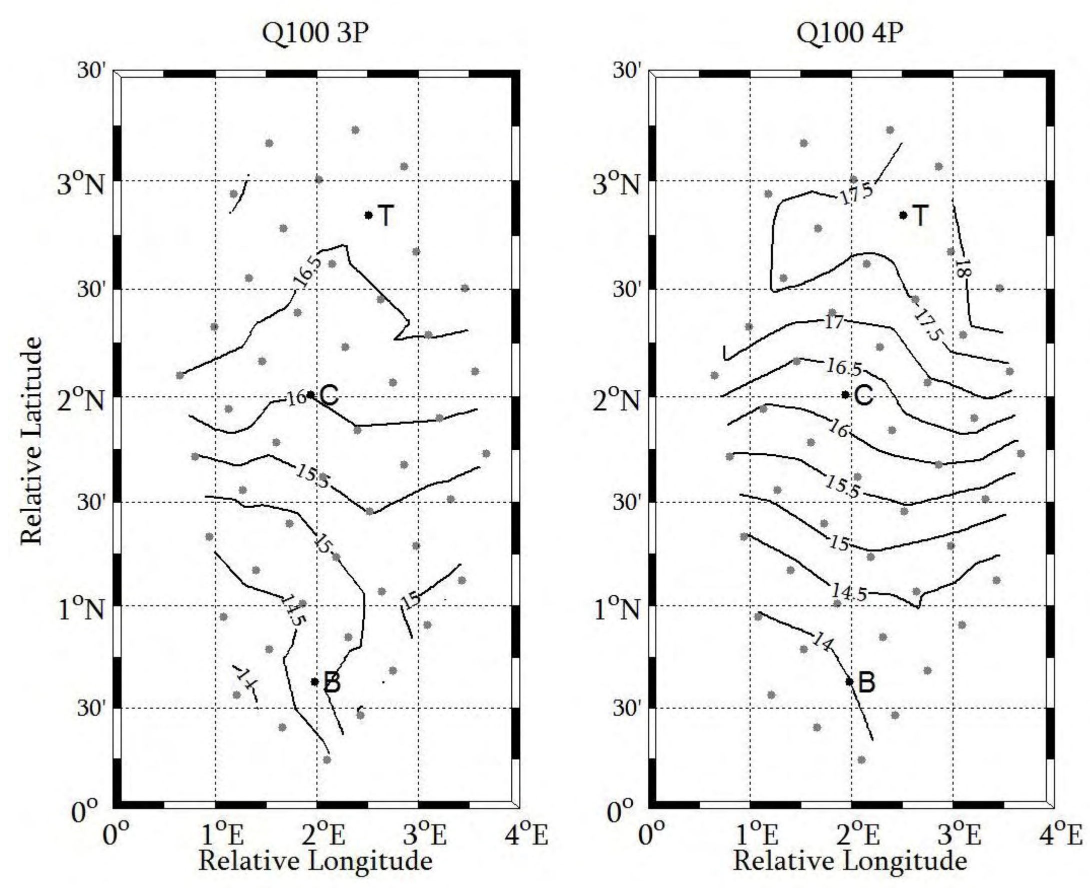

Figure4: Q_{100} median return values for the 3 and 4P models, as contours for all 50 locations (with relative longitude and latitude) for 75% quantile and 48 hour blocks. The contours show that the Q_{100} median return level increases towards the deeper, more exposed northern sites. Locations T, C & B are indicated.

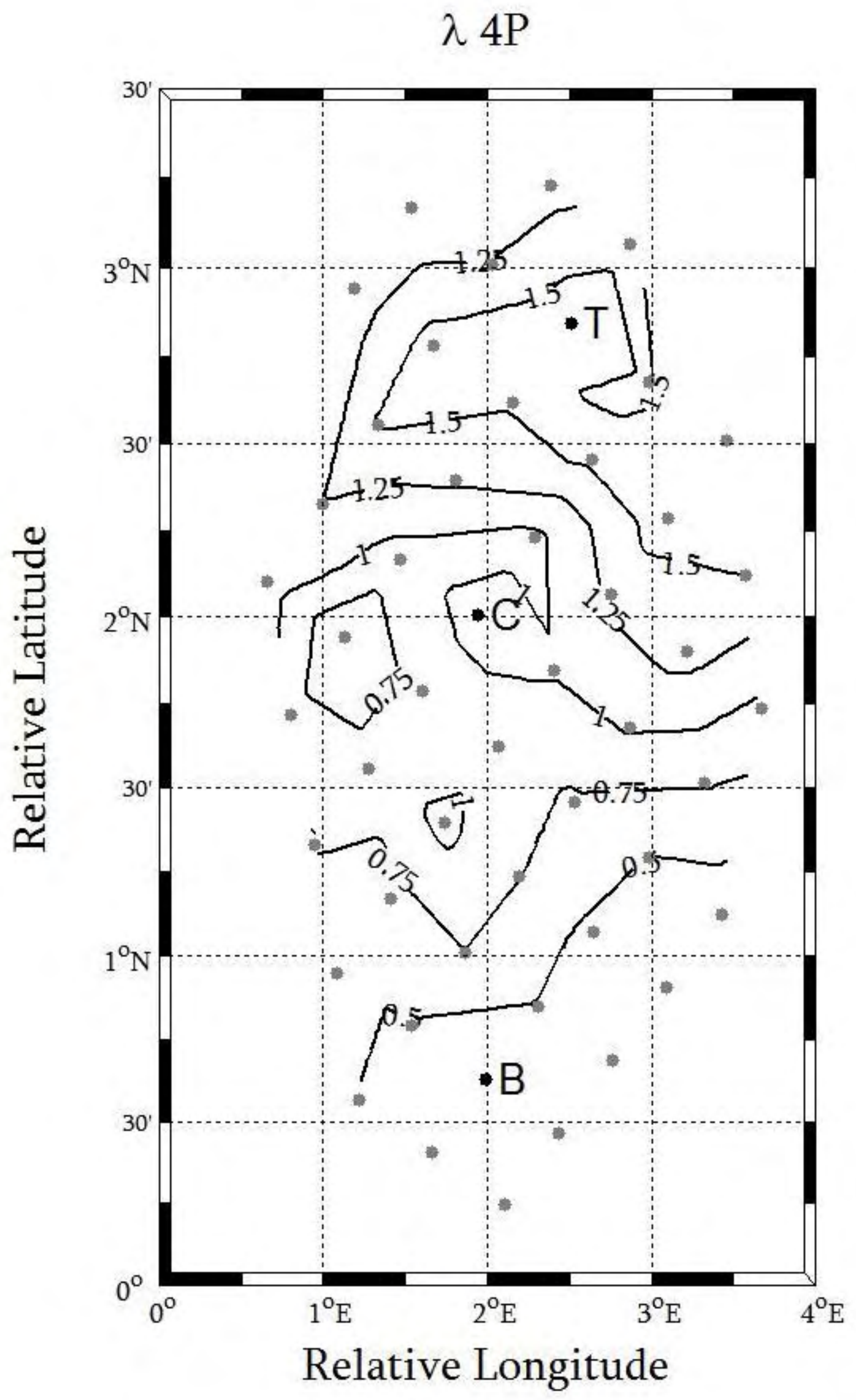

Figure5: Median values of scale parameter λ from 4P model as contours over locations. The value of λ varies from approximately 0.4 – 1.8 and increases moving northwards into deeper water. This indicates data scale transformations from the square root to the square of the data may be appropriate to enhance t to the extreme value model and that this decision is linked to location. (Locations T, C & B are indicated)

Discussion and Conclusions

- Results in figure 5 (above) suggest that a range of values are indicated by the data and that the choice of measurement scale has a spatial link.
- The most northerly sites in the analysed region, which are deeper and more exposed requiring a linear or squared measurement scale.
- By contrast the more sheltered sites require a linear or square root scale in order to maintain appropriate tail behaviour
- By modelling the data with an extra scale parameter λ we take account of this uncertainty and give more robust return level estimates.

Further Study

- We plan to model the whole of the North Sea to examine effect over a large region.
- Simultaneous modelling of covariate effects and scale effects would allow for use of more data and more robust results.

References

- G. E. P. Box and D. R. Cox. An analysis of transformations. Journal of the Royal Statistics Society. Series B (Methodological), 26:211-252, 1964.
- S. Coles. An introduction to statistical modelling of extreme values. Springer, London, 2001.
- J. L. Wadsworth, J. A. Tawn, and P. Jonathan. Accounting for choice of measurement scale in extreme value modelling. Annals of Applied Statistics, 4:1558{-578, 2010.